COMENTARIS DE LLIBRES

TRATAMIENTO ESTADÍSTICO DE DATOS
MÉTODOS Y PROGRAMAS

Autores: L.Lebart, A.Morineau, J.P.Fénelon

Editorial: Marcombo, Barcelona

520 + XI páginas, anexos y apéndices.

La obra que reseñamos es traducción de "Traitement des doneés statistiques. Méthodes et programmes", Bordas, Paris. Vista en conjunto, constituye una buena síntesis de las técnicas clásicas y recientes de la estadística y la informática. Como advierten los autores, el lector de "Tratamiento estadístico de datos" debe tener conocimientos de cálculo de probabilidades, estadística aplicada e informática (a nivel de conocer uno o varios lenguajes de programación). Por lo tanto se trata de una obra para profesionales o usuarios conscientes de la estadística, o para estudiantes de un segundo ciclo de estadística aplicada.

La obra incorpora las técnicas recientes de la estadística (análisis de datos multidimensionales, análisis exploratorio de datos, estimador jacknife, clasificación automática, etc.), junto con otras técnicas clásicas (pruebas no paramétricas, análisis de la varianza, etc.). Dividida en cinco grandes capítulos, en el capítulo **1** se presentan los aspectos esenciales de la teoría de la probabilidad, en forma condensada pero bastante completa (probabilidad en álgebras de sucesos, variables aleatorias finitas y generales, distribuciones bivariantes y multivariantes, teoremas límites), junto con algunas nociones de estadística (muestras, estadísticos, estimadores) y aplicaciones de las mismas (métodos de Montecarlo, simulación, método "Jacknife").

La introducción al razonamiento estadístico se hace por la via no paramétrica, contrariamente a lo que sería usual. Así, los métodos no paramétricos, de uso más general y explicación relativamente sencilla, son debidamente explicados en el capítulo 2. La exposición es bastante didáctica y rica en ejemplos, Además de las pruebas clásicas (de los signos, de Wilcoxon-Mann-Whitney, binomiales) se presentan otras menos habituales en la literatura sobre al tema.

Las técnicas de regresión y análisis de la varianza se explican en el capítulo 3, como una consecuencia del modelo lineal, que se define y estudian sus propiedades, primero sin hipótesis de normalidad y a continuación suponiendo normalidad sobre la variable dependiente o endógena. En el primer caso se estudia la estimación de los parámetros del modelo y en el segundo caso se llega al test F y sus aplicaciones. Todo ello permite enfocar i con comodidad el análisis de la varianza (de un criterio o varios) y el análisis de la covarianza. Es de destacar que el clásico test t de Student no se explica como tal, y que sólo se aborda el modelo lineal con matriz de diseño de rango máximo, lo que permite evitar las matrices inversas generalizadas, pero impide abordar el caso general.

El análisis de datos, en sentido multidimensional, es decir, el tratamiento estadístico simultáneo de varias variables, se inicia en el capítulo 4. En realidad es a partir de aquí donde los autores aportan sus contribuciones más originales al tema, pues son acreditados especialistas. Distinguen (muy acertadamente) entre los "métodos factoriales" o representación a través de modelos continuos, casi siempre a lo largo de uno o dos ejes de coordenadas, y los "métodos de clasificación" que consiste en agrupar los datos a clasificar en clases o familias de clases, y que en algunos casos se representen mediante grafos. Los "métodos factoriales" son objeto del capítulo 4.

Los autores (muy influidos por las ideas subyacentes en el análisis de correspondencias) inician el tema con un ajuste de una matriz n » p de datos (n filas y p columnas) en el espacio R^p (representación de n filas a través de p columnas), y en el espacio R^n (problema inverso), para, seguidamente, relacionar ambos tipos de representaciones.

Este enfoque es bastante interesante, pues permite abordar, bajo un punto de vista común, el análisis de componentes principales, el análisis en componentes principales normalizados, el análisis de los rangos, el análisis de las correlaciones parciales y el análisis de correspondencias. Todos estos métodos son explicados con detalle y debidamente ilustrados con ejemplos. Seguidamente en el mismo capítulo, pero formando una sección independiente de los anteriores métodos (pues obedecen a un esquema teórico diferente) se hace una breve exposición del análisis factorial, la regresión ortogonal, la regresión sobre componentes principales, el análisis canónico (o de correlación canónica) y el análisis discriminante, aunque sin ejemplos ilustrativos. No se aborda el análisis de coordenadas principales.

La clasificación automática (capítulo 5) se expone de una manera concisa pero suficientemente clara. La clasificación ascendente jerárquica se plantea a través de la noción de distancia ultramétrica y su relación con el concepto de jerarquía indexada. Se describe entonces con detalle la construcción de la métrica subdominante, también llamado método del mínimo. A continuación se describen algunos algoritmos de construcción de grafos parciales mínimos, es decir, grafos conexos sin ciclos cuyos vértices son los objetos a representar, y que reciben el nombre de árboles. En este apartado se encuentra a faltar una referencia a la desigualdad aditiva, también llamado axioma de los cuatro puntos, generalización de la desigualdad ultramétrica, que darían soporte teórico a las representaciones mediante árboles. Tras tratar algunos otros aspectos, se describe la clasificación no jerárquica, con especial énfasis en el método de clasificación alrededor de centros móviles, sobre el cual los autores hacen algunas observaciones respecto a la obtención de óptimos locales, y como resolver, aunque parcialmente, el problema de encontrar particiones óptimas. La literatura citada sobre el tema no es muy reciente.

La obra contiene varios apéndices. El primero se dedica a presentar el lenguaje FORTRAN y una relación de los principales paquetes de programas estadísticos. El segundo es una breve presentación del lenguaje APL y el tercero es un repaso de los principales conceptos de cálculo matricial. Contiene también las correspondientes tablas estadísticas.

Por otra parte, cada capítulo contiene un anexo con diversos programas escritos en FORTRAN y APL, con la finalidad de proporcionar al lector la facilidad de poder tratar informáticamente el material estadístico de

acuerdo con las técnicas explicadas previamente. Se trata de una biblioteca de programas modular, con finalidad pedagógica, pensada para que pueda ser adaptada por el usuario a sus propias necesidades. Contiene algunos programas del SPAD (Système portable pour L'Analyse des Données), y otros programas propios (pruebas fisherianas, modelo lineal, histogramas). No contiene correspondencias múltiples. La utilización de todos los programas está debidamente explicada.

La traducción es correcta. Algunos términos informáticos franceses son presentados en su versión inglesa (software por logiciel, etc.). Esta versión castellana incluye también algunos párrafos complementarios que se añaden al original en francés.

C.M. Cuadras